

Incomplete Polynomial Diagonals-Parameter Symmetry Model and Decomposition of Incomplete Symmetry Model for Square Contingency Tables

Hiroyuki Kurakami^{*1}, Akihiro Fujimura², Sadao Tomizawa³

^{*1,2}Graduate School of Science and Technology, Tokyo University of Science, ³Faculty of Science and Technology, Tokyo University of Science

^{*1,2,3}Noda City, Chiba, 278-8510, Japan

^{*1}h-kurakami@mti.biglobe.ne.jp; ²ak74i.hlww.ocjonrail@gmail.com; ³tomizawa@is.noda.tus.ac.jp

Received 13 December, 2013; Revised 10 March, 2014; Accepted 20 March, 2014; Published 18 May, 2014

© 2014 Science and Engineering Publishing Company

Abstract

For square contingency tables with ordered categories, Tomizawa (1990) considered the polynomial diagonals-parameter symmetry (PDPS) model. The present paper proposes the incomplete PDPS model which has the structure of PDPS for the partial cells of off-diagonal cells except a specified pair of cells (u, v) and (v, u) , $u \neq v$, in the table. It also gives the decomposition of incomplete symmetry model into the incomplete simple PDPS model and the incomplete mean equality model. An example is given.

Keywords

Incomplete Symmetry Model; Polynomial Diagonals-Parameter Symmetry; Square Contingency Table; Symmetry

Introduction

For an $r \times r$ square contingency table with the same ordinal row and column classifications, let p_{ij} denote the probability that an observation will fall in the i th row and j th column of the table ($i = 1, \dots, r; j = 1, \dots, r$). Goodman (1979) considered the diagonals-parameter symmetry (DPS) model defined by

$$p_{ij} = \begin{cases} \theta_{j-i} \phi_{ij} & (i < j), \\ \phi_{ij} & (i \geq j), \end{cases}$$

where $\phi_{ij} = \phi_{ji}$. This model states that the probability that an observation will fall in a cell (i, j) for $i < j$ is θ_{j-i} times higher than the probability that the

observation falls in the cell (j, i) . Special cases of this model obtained by putting $\theta_1 = \dots = \theta_{r-1} = 1$ and $\theta_1 = \dots = \theta_{r-1} (= \theta)$ are the symmetry (S) model (Bowker, 1948), and the conditional symmetry (CS) model (McCullagh, 1978), respectively.

Agresti (1983) considered the linear diagonals-parameter symmetry (LDPS) model defined by

$$p_{ij} = \begin{cases} \theta^{j-i} \phi_{ij} & (i < j), \\ \phi_{ij} & (i \geq j), \end{cases}$$

where $\phi_{ij} = \phi_{ji}$. This model is a special case of the DPS model.

Tomizawa (1987) considered the 2-ratios-parameter symmetry (2RPS) model defined by

$$p_{ij} = \begin{cases} \varphi \theta^{j-i} \phi_{ij} & (i < j), \\ \phi_{ij} & (i \geq j), \end{cases}$$

where $\phi_{ij} = \phi_{ji}$. Special cases of this model obtained by putting $\theta = 1$ and $\varphi = 1$ are the CS and LDPS models, respectively.

Tomizawa (1990) considered the polynomial diagonals-parameter symmetry (PDPS) model defined by

$$p_{ij} = \begin{cases} \phi_{ij} \prod_{k=0}^{r-2} \theta_k^{(j-i)^k} & (i < j), \\ \phi_{ij} & (i \geq j), \end{cases}$$

where $\phi_{ij} = \phi_{ji}$. This model is a generalization of the S, CS, LDPS and 2RPS models and another expression of the DPS model, see Tomizawa (1990). Special cases of this model obtained by putting $\theta_0 = \theta_1 = \dots = \theta_{r-2} = 1$, $\theta_1 = \theta_2 = \dots = \theta_{r-2} = 1$, $\theta_0 = \theta_2 = \dots = \theta_{r-2} = 1$ and $\theta_2 = \theta_3 = \dots = \theta_{r-2} = 1$ are the S, CS, LDPS and 2RPS models, respectively.

For analyzing the data in square tables, when certain model does not hold, we are interested in finding which cell influences the lack of the structure of the model. The incomplete S and incomplete CS models are considered by Tomizawa and Tokunaga (2006) and the incomplete DPS model is considered by Kurakami, Fujimura and Tomizawa (2013) which have a structure of the corresponding S, CS and DPS, respectively, except a specified pair of cells (u, v) and (v, u) , where $1 \leq u < v \leq r$.

The present paper proposes the incomplete PDPS model (and the incomplete LDPS and 2RPS models) which has the structure of PDPS for the partial cells of off-diagonals cells in the table. It also gives the decomposition of the incomplete S model into the incomplete LDPS model and the incomplete mean equality model.

Incomplete Polynomial Diagonals-Parameter Symmetry Model

Consider the incomplete PDPS model as follows; for a fixed (u, v) , where $1 \leq u < v \leq r$,

$$p_{ij} = \begin{cases} \phi_{ij} \prod_{k=0}^{r-2} \theta_k^{(j-i)^k} & (i < j, (i, j) \neq (u, v)), \\ \phi_{ij} & (i \geq j, (i, j) \neq (v, u)), \end{cases}$$

where $\phi_{ij} = \phi_{ji}$. This model has the structure of PDPS for the partial cells of off-diagonal cells except a specified pair of cells (u, v) and (v, u) in the table. We denote this model by PDPS(u, v). This model states that the probability that an observation will fall in a cell (i, j) for $i < j$ except the pair of cells (u, v) and (v, u) is $\prod_{k=0}^{r-2} \theta_k^{(j-i)^k}$ times higher than the probability that it falls in the cell (j, i) . Note that the PDPS model implies the PDPS(u, v) model. For analyzing the data, when the PDPS model does not hold, it may be possible to find which cell influences the lack of the structure of PDPS by applying various PDPS(u, v) models. Special cases of the PDPS(u, v) model obtained

by putting $\theta_0 = \theta_1 = \dots = \theta_{r-2} = 1$, $\theta_1 = \theta_2 = \dots = \theta_{r-2} = 1$, $\theta_0 = \theta_2 = \dots = \theta_{r-2} = 1$ and $\theta_2 = \theta_3 = \dots = \theta_{r-2} = 1$ are the S(u, v), CS(u, v), LDPS(u, v) and 2RPS(u, v) models, respectively. The LDPS(u, v) and 2RPS(u, v) models indicate the structure of the incomplete LDPS and incomplete 2RPS, respectively. Note that the LDPS(u, v) and 2RPS(u, v) models are new models. The LDPS(u, v) model is defined by

$$p_{ij} = \begin{cases} \theta^{j-i} \phi_{ij} & (i < j, (i, j) \neq (u, v)), \\ \phi_{ij} & (i \geq j, (i, j) \neq (v, u)), \end{cases}$$

where $\phi_{ij} = \phi_{ji}$. A special case of this model obtained by putting $\theta = 1$ is the S(u, v) model. The 2RPS(u, v) model is defined by

$$p_{ij} = \begin{cases} \phi \theta^{j-i} \phi_{ij} & (i < j, (i, j) \neq (u, v)), \\ \phi_{ij} & (i \geq j, (i, j) \neq (v, u)), \end{cases}$$

where $\phi_{ij} = \phi_{ji}$. Special cases of this model obtained by putting $\theta = 1$ and $\phi = 1$ are the CS(u, v) and LDPS(u, v) models, respectively. We point out that the PDPS(u, v) with no restriction between parameters $\{\theta_k\}$ is equivalent to the DPS(u, v) because the PDPS with no restriction is equivalent to the DPS.

Decomposition of Incomplete Symmetry Model

Let X and Y denote the row and column variables, respectively. Tahata, Yamamoto and Tomizawa (2008), and Tahata, Yamamoto and Tomizawa (2013) gave the decomposition of the S model into the LDPS and the mean equality (ME) model. The ME model is defined by $E(X) = E(Y)$. For a fixed (u, v) , where $1 \leq u < v \leq r$, we define the incomplete ME (ME(u, v)) model by $E(X | (X, Y) \neq (u, v) \text{ or } (v, u)) = E(Y | (X, Y) \neq (u, v) \text{ or } (v, u))$. We obtain the following theorem.

Theorem 1. For a fixed (u, v) , where $1 \leq u < v \leq r$, the S(u, v) model holds if and only if both the LDPS(u, v) and ME(u, v) models hold.

Proof. Let

$$q_{ij} = \frac{1}{c} p_{ij} \quad ((i, j) \neq (u, v), (v, u)),$$

$$c = \sum_{(s,t) \neq (u,v), (v,u)} p_{st}.$$

The S(u, v) model is expressed as

TABLE 1 NUMBERS OF DEGREES OF FREEDOM FOR MODELS.

Models	Degrees of freedom
S	$r(r-1)/2$
CS	$(r+1)(r-2)/2$
PDPS	$(r-1)(r-2)/2$
LDPS	$(r+1)(r-2)/2$
2RPS	$(r^2-r-4)/2$
$S(u,v)$	$(r+1)(r-2)/2$
$CS(u,v)$	$(r^2-r-4)/2$
$PDPS(u,v)$	$r(r-3)/2$
$LDPS(u,v)$	$(r^2-r-4)/2$
$2RPS(u,v)$	$(r+2)(r-3)/2$
$ME(u,v)$	1

$$q_{ij} = q_{ji} \quad ((i, j) \neq (u, v), (v, u)).$$

The $ME(u, v)$ model is expressed as

$$\sum_{(i,j) \neq (u,v), (v,u)} \sum (iq_{ij} - jq_{ij}) = 0.$$

This may be expressed as

$$\sum_{\substack{i < j \\ (i,j) \neq (u,v)}} \sum (j-i)(q_{ij} - q_{ji}) = 0. \quad (1)$$

If the $S(u, v)$ model holds, then the $LDPS(u, v)$ and the $ME(u, v)$ models hold. Conversely, if the $LDPS(u, v)$ and the $ME(u, v)$ models hold, then the equation (1) is expressed as

$$\sum_{\substack{i < j \\ (i,j) \neq (u,v)}} \sum (j-i)(\theta^{j-i} - 1)\phi_{ji} = 0,$$

where ϕ_{ij} satisfies $\phi_{ij} = \phi_{ji}$. Thus we obtain $\theta = 1$. Namely the $S(u, v)$ model holds. The proof is completed.

Goodness-of-fit Test

Assume that a multinomial distribution is applied to the $r \times r$ table. The maximum likelihood estimates (MLEs) of expected frequencies under the incomplete PDPS model could be obtained using an iterative procedure, for example, the Newton-Raphson method in the log-likelihood equations. Let n_{ij} denote the observed frequency in the i th row and j th column of the $r \times r$ table. Let m_{ij} denote the corresponding expected frequency ($i=1, \dots, r; j=1, \dots, r$) and let $\hat{m}_{ij}$ denote the MLE of m_{ij} under the model. The likelihood ratio chi-squared statistic for testing the goodness-of-fit of the model is

$$G^2 = 2 \sum_{i=1}^r \sum_{j=1}^r n_{ij} \log \left(\frac{n_{ij}}{\hat{m}_{ij}} \right).$$

TABLE 2 MOTHER'S EDUCATION BY FATHER'S EDUCATION; FROM MULLINS AND SITES (1984) FOR A SAMPLE OF EMINENT BLACK AMERICANS. THE PARENTHEZIZED VALUES ARE THE MLES OF THE CS(1, 3) MODEL.

Mother's Education	Father's Education				Total
	(1)	(2)	(3)	(4)	
(1)	81 (81.00)	3 (6.71)	9 (9.00)	11 (12.64)	104
(2)	14 (10.29)	8 (8.00)	9 (6.32)	6 (4.74)	37
(3)	43 (43.00)	7 (9.68)	43 (43.00)	18 (16.59)	111
(4)	21 (19.36)	6 (7.26)	24 (25.41)	87 (87.00)	138
Total	159	24	85	122	390

NOTE: (1) 8TH GRADE OR LESS, (2) PART HIGH SCHOOL, (3) HIGH SCHOOL, (4) COLLEGE

The numbers of degrees of freedom for various PDPS and PDPS(u, v) models are described in Table 1.

Example

The data in Table 2, taken directly from Mullins and Sites (1984) for a sample of eminent black Americans defined as persons having biographical sketch in the publication *Who's Who Among Black Americans*, describes the cross-tabulating the mother with the father on educational attainment.

We see from Table 3 that the S, CS, PDPS, LDPS and 2RPS models fit these data poorly. So, we shall apply various incomplete models. We see from Table 3 that the CS(1, 3), PDPS(1, 3), LDPS(1, 3), 2RPS(1, 3), PDPS(1, 2) and PDPS(2, 4) models fit these data well, however, the other models fit these data poorly. Therefore, for example, we can see that the poor fit of the CS model (i.e., the PDPS model with $\theta_1 = \theta_2 = 1$) is caused by the lack of the structure of CS for the pair of cells (1, 3) and (3, 1). Under the CS(1, 3) model, the MLE of θ_0 is 0.65 and the MLE of p_{13}/p_{31} is 0.21. Note that the approximate 95% confidence interval of θ_0 is [0.41, 0.89], with the standard error 0.12. Therefore, under this model, the probability that an individual's mother's educational attainment is i and his/her father's attainment is $j (> i)$, where $(i, j) \neq (1, 3)$, is estimated to be 0.65 times higher than the probability that an individual's mother's attainment is j and his/her father's attainment is i , and the probability that an individual's mother's attainment is (1) and his/her father's attainment is (3) is estimated to be 0.21 times higher than the probability that an individual's mother's attainment is (3) and his/her father's attainment is (1).

TABLE 3 VALUES OF G^2 APPLIED TO THE DATA IN TABLE2.

Models	Degrees of freedom	G^2
S	6	36.18*
CS	5	15.40*
PDPS	3	10.96*
LDPS	5	14.82*
2RPS	4	14.25*
S(1, 2)	5	28.46*
S(1, 3)	5	12.01*
S(1, 4)	5	33.00*
S(2, 3)	5	35.93*
S(2, 4)	5	36.18*
S(3, 4)	5	35.32*
CS(1, 2)	4	13.25*
CS(1, 3)	4	6.72
CS(1, 4)	4	15.35*
CS(2, 3)	4	11.26*
CS(2, 4)	4	13.76*
CS(3, 4)	4	12.89*
PDPS(1, 2)	2	5.98
PDPS(1, 3)	2	5.81
PDPS(2, 3)	2	8.76*
PDPS(2, 4)	2	5.81
PDPS(3, 4)	2	10.63*
LDPS(1, 2)	4	10.64*
LDPS(1, 3)	4	7.01
LDPS(1, 4)	4	11.41*
LDPS(2, 3)	4	13.12*
LDPS(2, 4)	4	12.85*
LDPS(3, 4)	4	14.70*
2RPS(1, 2)	3	10.64*
2RPS(1, 3)	3	6.53
2RPS(1, 4)	3	10.96*
2RPS(2, 3)	3	11.08*
2RPS(2, 4)	3	12.27*
2RPS(3, 4)	3	12.85*
ME(1, 2)	1	17.52*
ME(1, 3)	1	4.97*
ME(1, 4)	1	21.77*
ME(2, 3)	1	22.04*
ME(2, 4)	1	22.49*
ME(3, 4)	1	19.99*

* means the 5% significant

Therefore, we can see that the pair of cells (1, 3) and (3, 1) influences the lack of the structure of CS. We point out that the CS model fits the data poorly but the CS(1, 3) model fit the data well, and so the CS(1, 3) model is useful for seeing the reason of the poor fit of the CS model as described above. Under the other models, similar explanations can be obtained although those are omitted.

The PDPS model fits these data poorly, however, each of PDPS(1, 2), PDPS(1, 3) and PDPS(2, 4) models fits these data well. Therefore the pair of cells (1, 2) and (2, 1) (or the pair of (1, 3) and (3, 1), or the pair of (2, 4) and (4, 2)) may influences the lack of the structure of PDPS.

The S(1, 3) and ME(1, 3) models fit these data poorly, however, the LDPS(1, 3) fits these data well. Therefore, it is seen from Theorem 1 that for these data, the poor fit of the S(1, 3) model is caused by the influence of the lack of structure of the ME(1, 3) model.

Conclusion

When the PDPS model fits the data poorly, the incomplete PDPS model (i.e., the PDPS(u, v) model) and Theorem 1 would be useful for finding which pair of cells influences the lack of the structure of PDPS (including the structure of S, CS, LDPS, 2RPS, and DPS).

ACKNOWLEDGMENT

The authors would like to thank the editor and anonymous reviewer for improving this paper.

REFERENCES

- Agresti, A. "A simple diagonals-parameter symmetry and quasi-symmetry model." *Statistics and Probability Letters*, vol. 1, (1983): 313-316.
- Bowker, A. H. "A test for symmetry in contingency tables." *Journal of the American Statistical Association*, vol. 43, (1948): 572-574.
- Goodman, L. A. "Multiplicative models for square contingency tables with ordered categories." *Biometrika*, vol. 66, (1979): 413-418.
- Kurakami, H., Fujimura, A. and Tomizawa, S. "Incomplete diagonals-parameter symmetry model for ordered square contingency tables." *Advance and Applications in Statistics*, vol. 33, (2013): 107-113.
- McCullagh, P. "A class of parametric models for the analysis of square contingency tables with ordered categories." *Biometrika*, vol. 65, (1978): 413-418.
- Mullins, E. J. and Sites, P. "The origins of contemporary eminent black Americans: A three-generation analysis of social origin." *American Sociological Review*, vol. 49, (1984): 672-685.
- Tahata, K., Yamamoto, H. and Tomizawa, S. "Orthogonality of decompositions of symmetry into extended symmetry and marginal equimoment for multi-way tables with ordered categories." *Austrian Journal of statistics*, vol. 37, (2008): 185-194.
- Tahata, K., Yamamoto, K. and Tomizawa, S. "Decomposition

- of symmetry model into three models for cumulative probabilities in square contingency tables." *European Journal of Pure and Applied Mathematics*, vol. 6, (2013): 299-306.
- Tomizawa, S. "Decompositions for 2-ratios-parameter symmetry model in square contingency tables with ordered categories." *Biometrical Journal*, vol. 29, (1987): 44-55.
- Tomizawa, S. "Polynomial diagonals-parameter symmetry model for square contingency tables with ordered categories." *Statistica*, vol. 50, (1990): 171-178.
- Tomizawa, S. and Tokunaga, S. "Incomplete symmetry and conditional symmetry models for square contingency tables with ordered categories." *Far East Journal of Theoretical Statistics*, vol. 19, (2006): 33-42.